

線型推測論

第03回 ガウス-マルコフの定理と最小二乗推定量(1)

2022/4/15

慶応義塾大学病院

長島 健悟

推定

- 統計学における推定
確率変数がある分布に従うとすると, その分布に含まれる**未知パラメータの真値**を**データから推測**すること
- モデル中の未知パラメータをデータ(観測値)に基づいて推測すること
- 点推定と区間推定があるので, まずは点推定に着目しよう

一般線型モデルの観測値の分布

- 実はモデル式は二組の確率変数を含む
- 今のところ分布などについては何も議論していない
- 一般線型モデルにおける分布の基本的な性質を確認する
- モデル式の仮定の確認

$$Y = X\beta + \epsilon$$

- 意味：Yが $X\beta$ と ϵ の和に分解できる

一般線型モデルの観測値の分布

- 誤差に対して追加の仮定を導入する
 - $\epsilon_1, \dots, \epsilon_n$ は互いに無相関で何らかの分布に従う確率変数
 - $E[\epsilon_i] = 0$
 - $\text{Var}[\epsilon_i] = \sigma^2 < \infty$, および無相関なので $\text{Cov}[\epsilon_i, \epsilon_j] = 0 \ (i \neq j)$
- ※誤差分布の独立性を仮定する場合もあるが, これは無相関よりも強い仮定

一般線型モデルの観測値の分布

- この仮定のもとでは . . .
- $\mathbf{X}$ と $\boldsymbol{\beta}$ は定数で, ϵ は確率変数
- 誤差の仮定から $\mathbf{Y}$ も確率変数
- モデル式の仮定から等式 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$ が成立する

一般線型モデルの観測値の分布

- 以下は簡単に分かる
 - 系3-1
 - $E[Y] = E[X\beta + \epsilon] = X\beta$
 - 系3-2
 - $\text{Cov}[Y] = \text{Cov}[X\beta + \epsilon] = \text{Cov}[\epsilon] = \sigma^2 \mathbf{I}_n$

期待値と分散の性質の復習

- a を定数, Y, Z を確率変数とする
 - $E[a + Y] = a + E[Y]$
 - $E[aY] = aE[Y]$
 - $\text{Var}[Y] = E[(Y - E[Y])^2] = E[Y^2 - 2YE[Y] +$

確率変数ベクトルの期待値と分散

- $E[\mathbf{Y}] = \begin{pmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{pmatrix}$
- $\text{Cov}[\mathbf{Y}] = E[(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])']$
- $\text{Cov}[\mathbf{XY}] = \mathbf{XCov}[\mathbf{Y}]\mathbf{X}'$
- $\mathbf{X} : p \times n$ 次元定数行列, $\mathbf{Y} : n \times 1$ 次元確率変数ベクトル

練習問題

- 系3-1
 - $E[Y] = E[X\beta + \epsilon] = X\beta$
- 期待値の性質より
 - $E[X\beta + \epsilon] = X\beta + E[\epsilon]$
- 誤差の仮定から $E[\epsilon] = 0$
- よって, $E[X\beta + \epsilon] = X\beta + 0 = X\beta$

点推定量

- 確率変数がある分布に従うとすると, その分布に含まれる**未知パラメータの真値**を**データから推測**すること
- 点推定量はデータを表わす確率変数の関数になっている
- これを式で表わしてみると $T = g(Y)$ と書ける

よい点推定量？

- $T = g(\mathbf{Y})$ を構成する g は無限に考えられる
- 例えば $T = Y_1$ でも何かの推定量と言える
- 我々は g の中でも良いものを探したい
 - 点推定量の良さを議論するには基準が必要

よい点推定量の基準

- 不偏性
 - 推定量の期待値が真値に一致
 - T が β の推定量であれば $E[T] = \beta$
- 最小分散性
 - 推定量の中で最も分散が小さい
- 最小分散不偏推定量
 - 不偏推定量の中で最も分散が小さい

線型推定量

- 以下では線型推定量を考える
- 推定量 $T = g(Y)$ が Y の線型式であるとき
この推定量を線型推定量とよぶ

- つまり, $c_1, \dots, c_n$ を定数として,

$$T = c_1 Y_1 + \dots + c_n Y_n$$

の形式の推定量

- 詳しい話は割愛するが, 一般線型モデル
では線型推定量が重要である

最良線型不偏推定量

- 線型不偏推定量
 - 不偏な線型推定量
- 最良線型不偏推定量 (Best Linear Unbiased Estimator; BLUE)
 - 線型不偏推定量の中で最も分散が小さい

例：秤量問題での点推定

- $Y_1 = -\beta_1 + \beta_2 + \epsilon_1$
- $Y_2 = \beta_1 + \beta_2 + \epsilon_2$
- $Y_3 = \beta_1 - \beta_2 + \epsilon_3$
- $E[\epsilon_i] = 0, \text{Var}[\epsilon_i] = \sigma^2, \text{Cov}[\epsilon_i, \epsilon_j] = 0$
(for $i \neq j$)
- 以下で β_1 のBLUEを求めてみよう

例：秤量問題での点推定

- 線型推定量を以下で定義する
 - $T = c_1 Y_1 + c_2 Y_2 + c_3 Y_3$
- c_i を定めれば, 一つの推定量が定まる

例：秤量問題での点推定

- まずは不偏性, $E[T] = \beta_1$, をみたす係数を探す
- $E[T] = E[c_1 Y_1 + c_2 Y_2 + c_3 Y_3] = c_1(-\beta_1 +$

例：秤量問題での点推定

- $-c_1 + c_2 + c_3 = 1$ かつ $c_1 + c_2 - c_3 = 0$ を満たす組み合わせも無数に存在する
- 次はこの中で最小分散のものを探そう

$$\text{Var}[T] = \text{Var}[c_1 Y_1 + c_2 Y_2 + c_3 Y_3] = c_1^2 \sigma^2 + c_2^2 \sigma^2 + c_3^2 \sigma^2 = (c_1^2 + c_2^2 + c_3^2) \sigma^2$$

- σ^2 は定数のため
- $c_1^2 + c_2^2 + c_3^2$ が最小のものを求めれば良い

例：秤量問題での点推定

- $-c_1 + c_2 + c_3 = 1$ かつ $c_1 + c_2 - c_3 = 0$ を満たし、 $c_1^2 + c_2^2 + c_3^2$ が最小になるものを求めれば良いことが分かる
- 何らかの制約条件下における最小化問題を解く方法があればよい
- Lagrangeの未定乗数法
等式制約下での最小化(最大化)問題の解放

Lagrangeの未定乗数法

- 等式制約 $h_1(\mathbf{x}) = 0, \dots, h_s(\mathbf{x}) = 0$ があるとし, 最小化したい関数を $f(\mathbf{x})$ とする
- $L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \lambda_1 h_1(\mathbf{x}) - \dots - \lambda_s h_s(\mathbf{x})$
- $\frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \mathbf{x}'} = 0, \frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}'} = 0$ の解は, 制約条件下での $f(\mathbf{x})$ の極値 (最小値または最大値) を与える
- ※ L をラグランジアンと呼ぶ

例：秤量問題での点推定

- 準備：等式制約に変形しておく
 - $-c_1 + c_2 + c_3 - 1 = 0, c_1 + c_2 - c_3 = 0$
- ラグランジアン L を構成する
 - $L(\mathbf{c}, \boldsymbol{\lambda}) = c_1^2 + c_2^2 + c_3^2 - \lambda_1(-c_1 + c_2 +$

例：秤量問題での点推定

- $\frac{\partial L(\mathbf{c}, \boldsymbol{\lambda})}{\partial c_1} = 2c_1 + \lambda_1 - \lambda_2$
- $\frac{\partial L(\mathbf{c}, \boldsymbol{\lambda})}{\partial c_2} = 2c_2 - \lambda_1 - \lambda_2$
- $\frac{\partial L(\mathbf{c}, \boldsymbol{\lambda})}{\partial c_3} = 2c_3 - \lambda_1 + \lambda_2$
- $\frac{\partial L(\mathbf{c}, \boldsymbol{\lambda})}{\partial \lambda_1} = c_1 - c_2 - c_3 + 1$
- $\frac{\partial L(\mathbf{c}, \boldsymbol{\lambda})}{\partial \lambda_2} = -c_1 - c_2 + c_3$

例：秤量問題での点推定

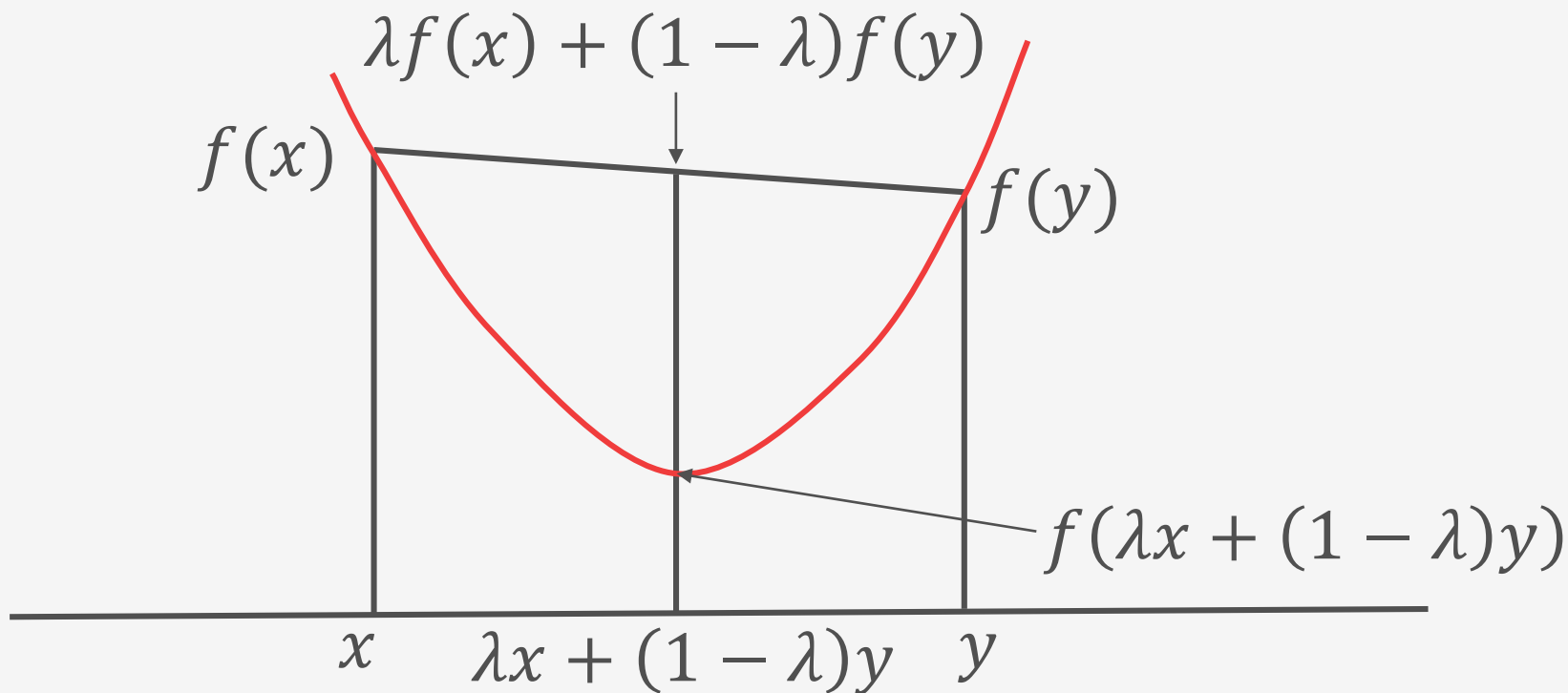
- 以下の連立方程式の解を求めればよい
 - $2c_1 + \lambda_1 - \lambda_2 = 0, 2c_2 - \lambda_1 - \lambda_2 = 0,$
 $2c_3 - \lambda_1 + \lambda_2 = 0, -c_1 + c_2 + c_3 - 1 = 0,$
 $c_1 + c_2 - c_3 = 0$
- 整理すると...
 - $c_1 + c_3 = 0, -c_1 + c_2 + c_3 = 1, c_1 + c_2 - c_3 = 0$
 - $c_1 = -\frac{1}{4}, c_2 = \frac{1}{2}, c_3 = \frac{1}{4} (\lambda_1 = \frac{3}{4}, \lambda_2 = \frac{1}{4})$

例：秤量問題での点推定

- 極値 $c_1 = -\frac{1}{4}$, $c_2 = \frac{1}{2}$, $c_3 = \frac{1}{4}$ が得られた
- $c_1^2 + c_2^2 + c_3^2$ は明らかに下に凸な関数なので, これは最小値である
- $T = -\frac{1}{4}Y_1 + \frac{1}{2}Y_2 + \frac{1}{4}Y_3$ は β_1 のBLUEである

凸関数の和は凸関数になる

- f が凸関数であることの定義
- 任意の二点 x, y と実数 $0 \leq \lambda \leq 1$ について
- $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$



凸関数の和が凸関数になる

- f も g も凸関数であるとする
- $$\begin{aligned} & f(\lambda x + (1 - \lambda)y) + g(\lambda x + (1 - \lambda)y) \\ & \leq \lambda f(x) + (1 - \lambda)f(y) + \lambda g(x) \\ & \quad + (1 - \lambda)g(y) \\ & = \lambda\{f(x) + g(x)\} + (1 - \lambda)\{f(y) + g(y)\} \end{aligned}$$
が成り立つから、 $h = f + g$ とおけば明らかに
- $h(\lambda x + (1 - \lambda)y) \leq \lambda h(x) + (1 - \lambda)h(y)$
- 凸関数の和が凸関数であることがわかる

まとめ

- BLUEは良い推定量と考えられる
- 秤量問題の例の様に, 線形推定量の制約付き最適化によってBLUEを求めることができた
- しかし, Lagrangeの未定乗数法で解くのは結構面倒である...
- もっとよい方法はないだろうか？